prevalence is low is, I believe, especially relevant when screening is not voluntary.

ACKNOWLEDGMENT

I thank Dr. Charles Hollingsworth of the Blood Resources Branch, Division of Blood Diseases and Resources, National Heart, Lung, and Blood Institute, for his information about ELISA and Western blot.

ADDITIONAL REFERENCES

GOLDBERG, J. D. and WITTES, J. T. (1981). The evaluation of medical screening procedures. *Amer. Statist.* **35** 4–11.

PEOPLE VS. COLLINS. (1968). *California Reporter* **66** 242–253. West Publishing Co., St. Paul, Minn.

SHAPIRO, S., STRAX, P., VENET, L. and VENET, W. (1973). Changes in five-year breast cancer mortality in a breast cancer screening program. *Seventh Nat. Cancer Conf. Proc.* 663–678.

Comment

Judith D. Goldberg

Professor Gastwirth addresses an important and interesting problem, the evaluation of medical screening procedures and programs.

The examples of AIDS screening with enzyme-linked immunosorbent assay (ELISA) and the use of the polygraph to detect deceptive individuals raise the broader questions of how to estimate the sensitivity and specificity of a screening test and how to implement and monitor the widespread use of a test in a population. The difficulties in the estimation of C arise from the practical issues of obtaining useful estimators of the sensitivity and specificity from incomplete data often in the absence of confirmatory testing of negatives on screening. The precision of $\hat{C}$ is a function of the precision of the component estimators that may themselves have large variances depending upon the method of estimation.

I would like to clarify the terminology used by Gastwirth. The traditional false negative rate or Neyman-Pearson type I error is defined as the proportion of "diseased" individuals who are negative on screen or (1 − sensitivity); the false positive rate, analogous to the type II error, is defined as the proportion of "nondiseased" individuals who are positive on screen or (1 − specificity) (Goldberg, 1975). These rates do not depend on the prevalence. The predictive value of a positive test, Gastwirth's C, the quantity of interest in this paper, was originally defined by Vecchio (1966) and does depend on the disease prevalence as does the predictive value of a negative test or $(1 - F)$ in Gastwirth's discussion. Gastwirth's F is not the traditional false negative rate nor is C the traditional true positive rate (Tables 1 to 3).

Goldberg and Wittes (1978) estimate the traditional false negative rate (1 − sensitivity) of a screening program, and not F as indicated by Gastwirth. The estimator depends on a capture-recapture estimator of the number of diseased individuals in the population. The observed data used to obtain the estimate are the numbers of positives on each of two distinct screening mechanisms; the prevalence of the disease is not estimated. The proposed estimators are useful when no confirmatory test is administered to individuals who are negative on the dual screening.

For example, in the Health Insurance Plan breast cancer screening program, a randomized trial designed to evaluate periodic screening with mammography and clinical examination, negatives on screen were returned to the population pool (Shapiro, Strax, Venet and Venet, 1973). The false negative rates estimated from this study vary and have wide confidence intervals even when the population is stratified into reasonably homogeneous groups.

In his paper, Gastwirth examines the estimated standard errors of $\hat{C}$ when prevalence and sample size vary with sensitivity and specificity held fixed (and assumed known or estimated from another source). Because $\hat{C}$ depends on the error rates as well, the sensitivity analysis should address the implications on estimation of the range of possible error rates for useful screening tests.

For diseases of low prevalence ($\pi \leq .05$), the bias in the estimator of the proportion positive on screening when there is misclassification depends primarily on the false positive rate (Goldberg, 1975). $\hat{C}$ depends on the misclassification rates both directly and indirectly through the estimator of the proportion positive on screening.

Gastwirth points out that prevalence can vary from group to group. It is just as likely that the false negative and false positive rates will vary from group to group for the same test (Goldberg, 1975). Thus, the analysis of the sensitivity of the precision of $\hat{C}$ should address first the sensitivity of $\hat{C}$ itself to underlying prevalence and error assumptions because a precise,

Judith D. Goldberg is Director, Statistical Design and Analysis, Medical Research Division, American Cyanamid Company, Pearl River, New York 10965.